BIO392
Bioinformatics of Genome Variations

Genomes: Core of "Personalized Health" & "Precision Medicine"

Michael Baudis **UZH SIB**
Computational Oncogenomics

University of Zurich[UZH]

Variant - Disease Knowledge Bases

VARIANT RESOURCES FOR CANCER GENOMICS

Resource name	Primary institute	Constituent Knowledge base	Cancer focused	Therapeutic evidence	Predisp. evidence	Diagnostic evidence	Prognostic evidence	Variant emphasis	URL
Cancer Genome Interpreter (CGI)	Institute for Research in Biomedicine, Barcelona, Spain	x	x	x				Somatic	https://www.cancergenomeinterpreter.org/home
Clinical Interpretation of Variants in Cancer (CIViC)	Washington University School of Medicine (WashU)	x	x	x	x	x	x	All variants	www.civicdb.org
JAX Clinical Knowledgebase (CKB)	The Jackson Laboratory	x	x	x	x	x	x	Somatic	https://ckb.jax.org/
Molecular Match	Molecular Match	x	x	x			x	Somatic	https://app.molecularmatch.com/
OncoKB	Memorial Sloan Kettering Cancer Center	x	x	x				Somatic	http://oncokb.org/#/
Precision Medicine Knowledgebase (PMKB)	Weill Cornell Medical College	x	x	x	x	x	x	Somatic	https://pmkb.weill.cornell.edu/
BRCA exchange	GA4GH	x	x		x			Germline	http://brcaexchange.org/
Cancer Driver Log (CanDL)	Ohio State University (OSU) / James Cancer Hospital		x	x				Somatic	https://candl.osu.edu/
Gene Drug Knowledge Database	Synapse		x	x		x	x	Somatic	https://www.synapse.org/#!Synapse:syn2370773/wiki/62707
MatchMiner	Dana-Farber Cancer Institute		x					Somatic	http://bcb.dfci.harvard.edu/knowledge-systems/
COSMIC Drug Resistance Curation	Wellcome Trust Sanger Institute		x	x				Somatic	http://cancer.sanger.ac.uk/cosmic/drug_resistance
My Cancer Genome	Vanderbilt University		x	x		x	x	Somatic	https://www.mycancergenome.org/
NCI Clinical Trials	National Cancer Institute of the National Institutes of Health		x					Somatic	www.cancer.gov/about-cancer/treatment/clinical-trials
Personalized Cancer Therapy Database	MD Anderson Cancer Center		x	x	x	x	x	Somatic	https://pct.mdanderson.org/#/home
ClinGen Knowledge Base	ClinGen				x			Germline	https://www.clinicalgenome.org/resources-tools/
ClinVar	National Center for Biotechnology Information			x	x			All variants	http://www.ncbi.nlm.nih.gov/clinvar/
Pharmacogenomics Knowledgebase (PharmGKB)	Stanford University			x				Germline	https://www.pharmgkb.org/
The Human Gene Mutation Database (HGMD)	Institute of Medical Genetics in Cardiff				x			Germline	http://www.hgmd.cf.ac.uk

Wagner et al. A harmonized meta-knowledgebase of clinical interpretations of cancer genomic variants. Nature Genetics (under review 2018)

CANCER VARIANT KNOWLEDGE BASES

- cancer variant knowledge databases report evidences for disease association (causative, therapeutic targets...)
- data selection is driven by arbitrary observations and sample selections
- limited overlap of reported variant associations is evidence for large gaps in knowledge

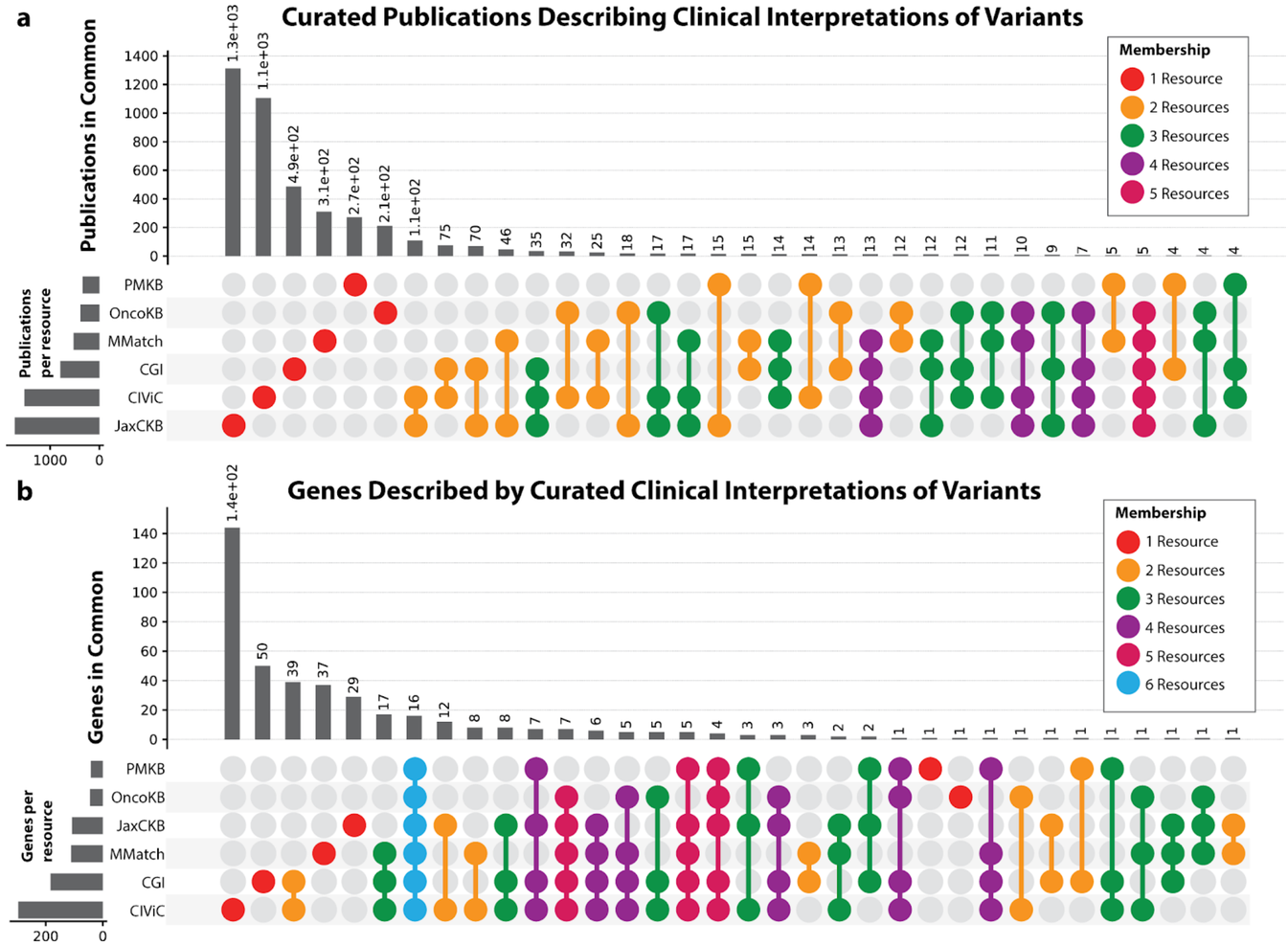


Figure S2 - Knowledgebase overlap

(a) Upset plot of publications supporting clinical interpretations of variants. the overwhelming majority of publications are observed in only 1 of 6 resources. **(b)** Upset plot of genes described by clinical interpretations of variants. Compared to other interpretation elements, genes are much more commonly shared between resources.

Wagner et al. A harmonized meta-knowledgebase of clinical interpretations of cancer genomic variants. Nature Genetics (under review 2018)

CANCER VARIANT KNOWLEDGE BASES

- cancer variant knowledge databases report evidences for disease association (causative, therapeutic targets...)
- data selection is driven by arbitrary observations and sample selections
- limited overlap of reported variant associations is evidence for large gaps in knowledge

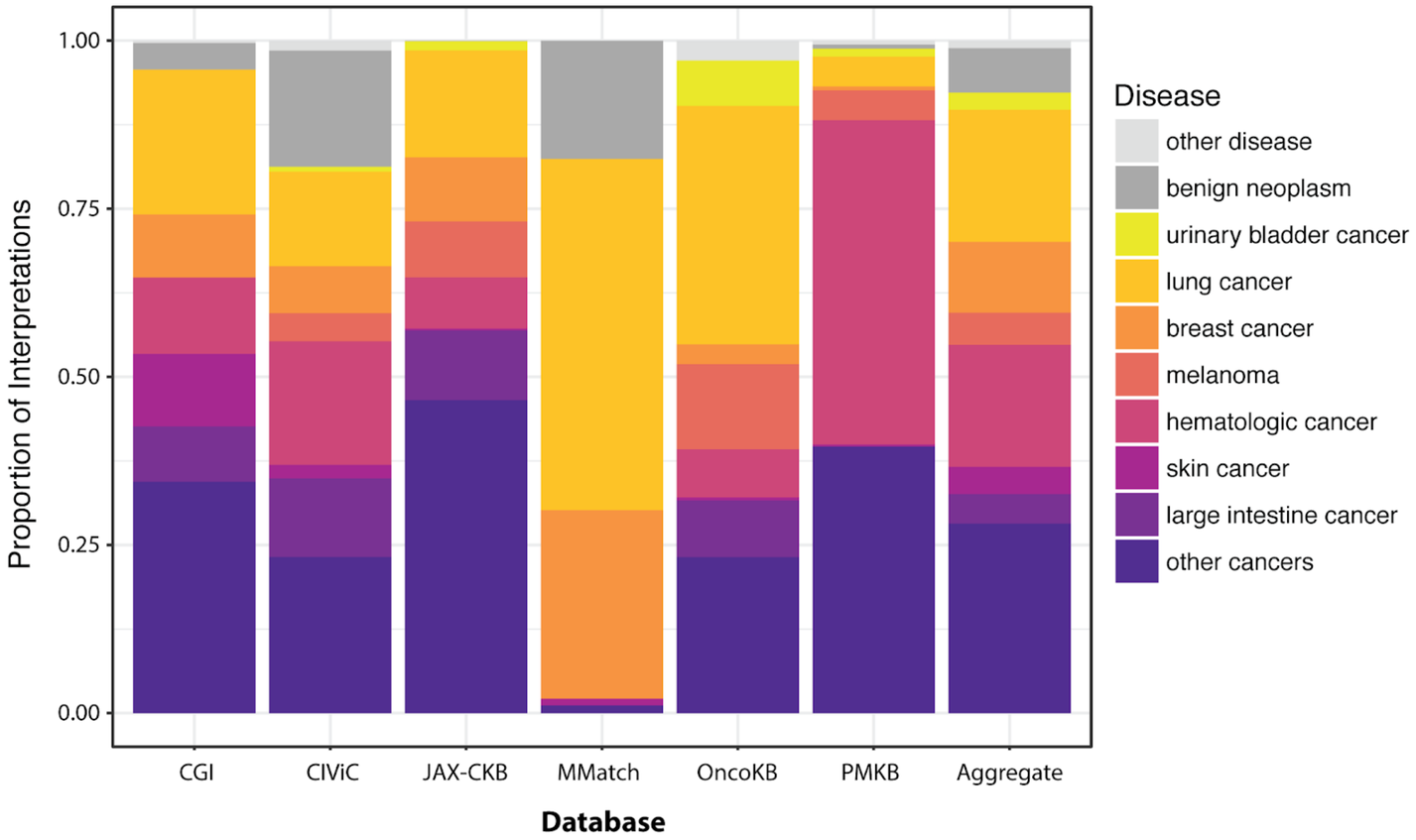


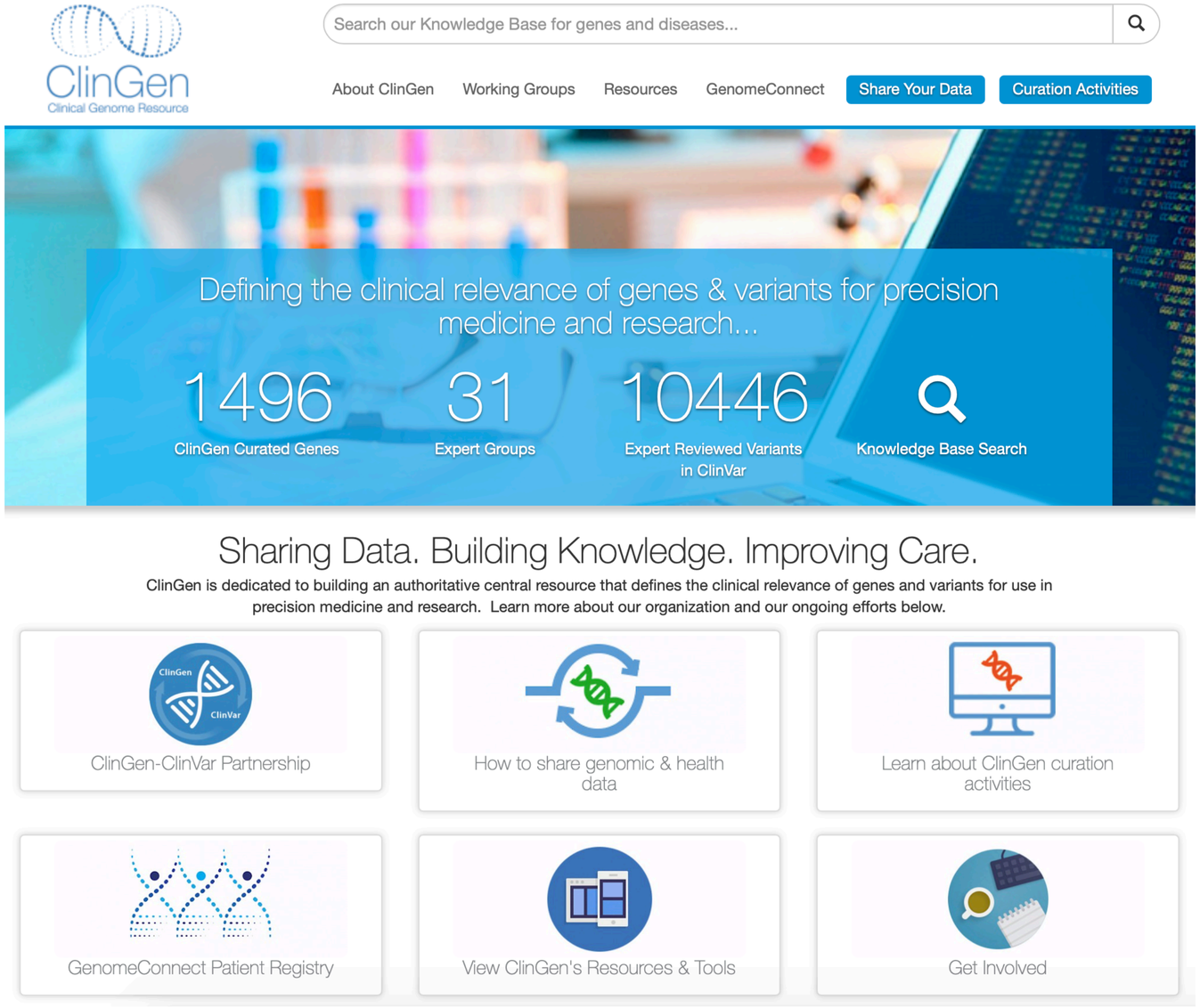
Figure S3 - Knowledgebase disease enrichment
Relative distribution of interpretations describing diseases across the VICC resources. Several resources are strongly enriched for one or more diseases compared to the entire dataset (see related **Table S8**).

Wagner et al. A harmonized meta-knowledgebase of clinical interpretations of cancer genomic variants. Nature Genetics (under review 2018)

RESOURCES FOR GENOMICS: CLINGEN

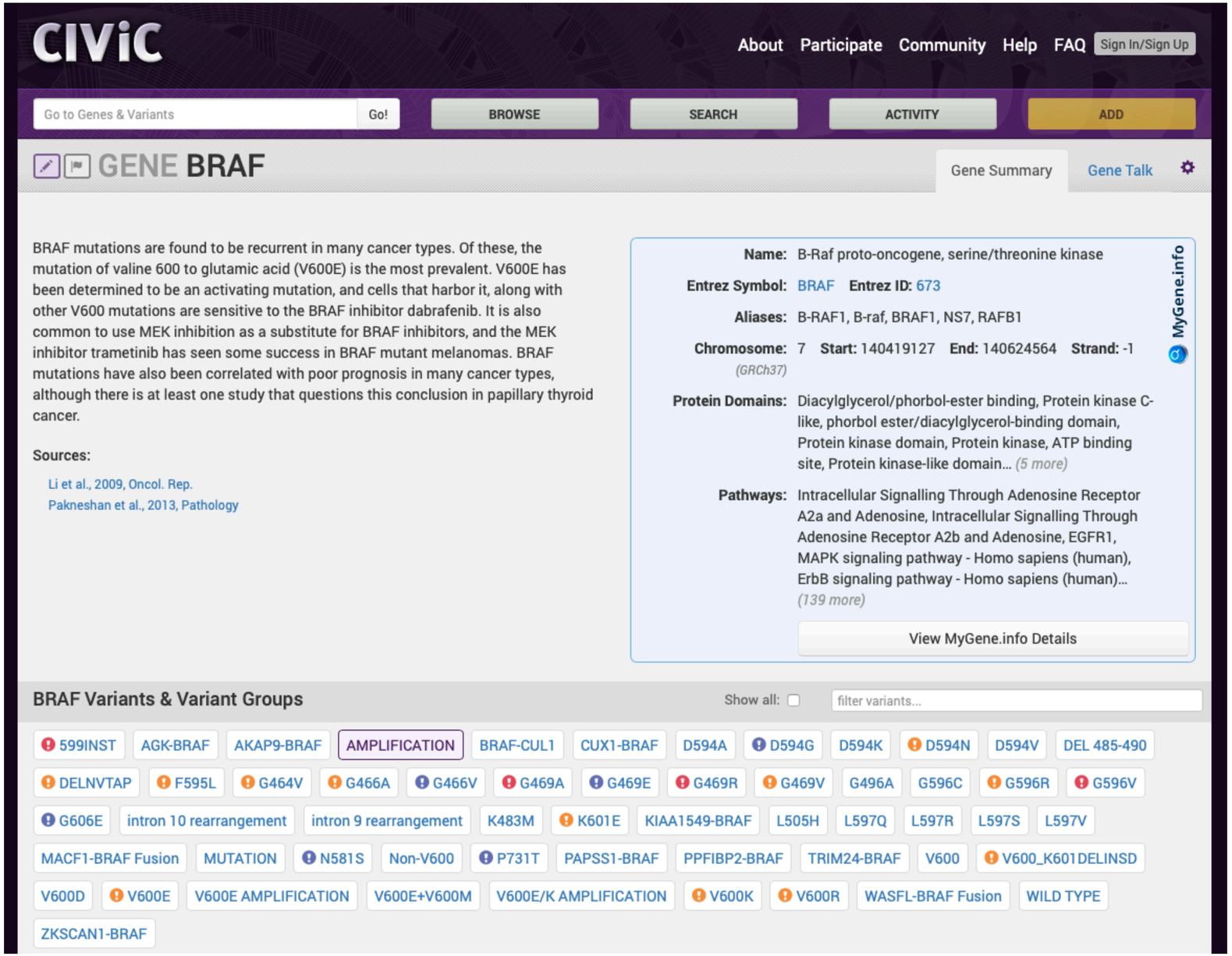
- "The Genomic Variant WG brings together representatives from the Sequence and Structural Variant communities for focused discussions on resolving discrepancies in variant interpretation and creating consistent curation guidelines."
- Interpreted genome variants with disease association

clinicalgenome.org



CANCER VARIANT KNOWLEDGE BASES: CIVIC

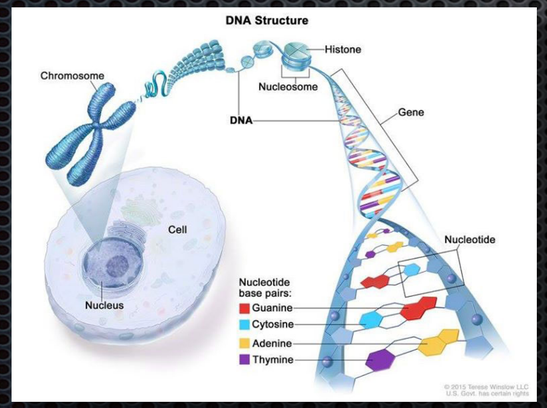
- "CIViC is a community-edited forum for discussion and interpretation of peer-reviewed publications pertaining to the clinical relevance of variants (or biomarker alterations) in cancer."

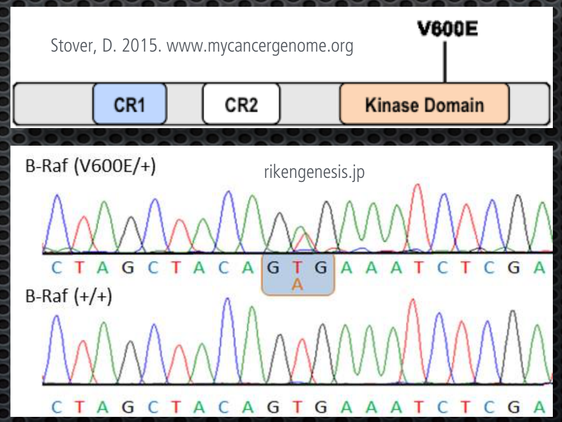


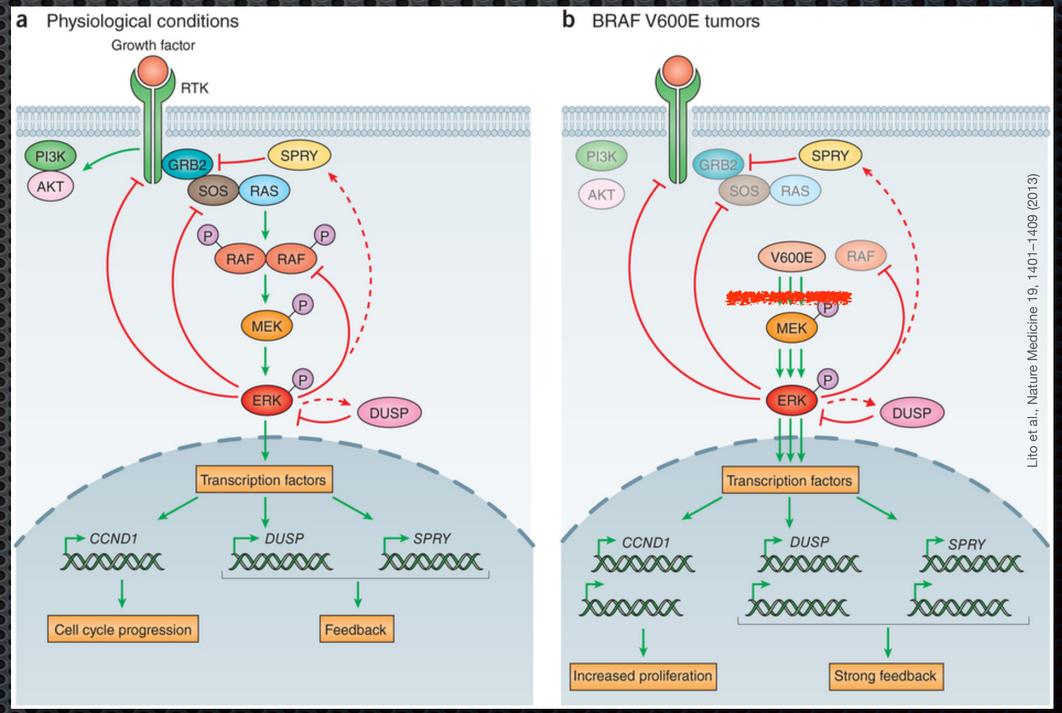
CIViC
Clinical Interpretations of Variants in Cancer

BRAF V600E (c.1799T>A) Mutation
Oncogene Activation by Single Nucleotide Alteration

- a single nucleotide exchange **T**hymidine > **A**denine leads to continuous RAF based activation of the MEK-ERK pathway
- BRAF V600E is a frequent mutation in >50% of malignant melanomas, but also CRC, lung ADC …
- pharmacologic block of B-Raf (e.g. through ***Vemurafenib***)







The BRAF V600E mutation leads to continuos phosphorylation of MEK, without the need for receptor based activation of the upstream pathway and loss of inhibitory feedback control.

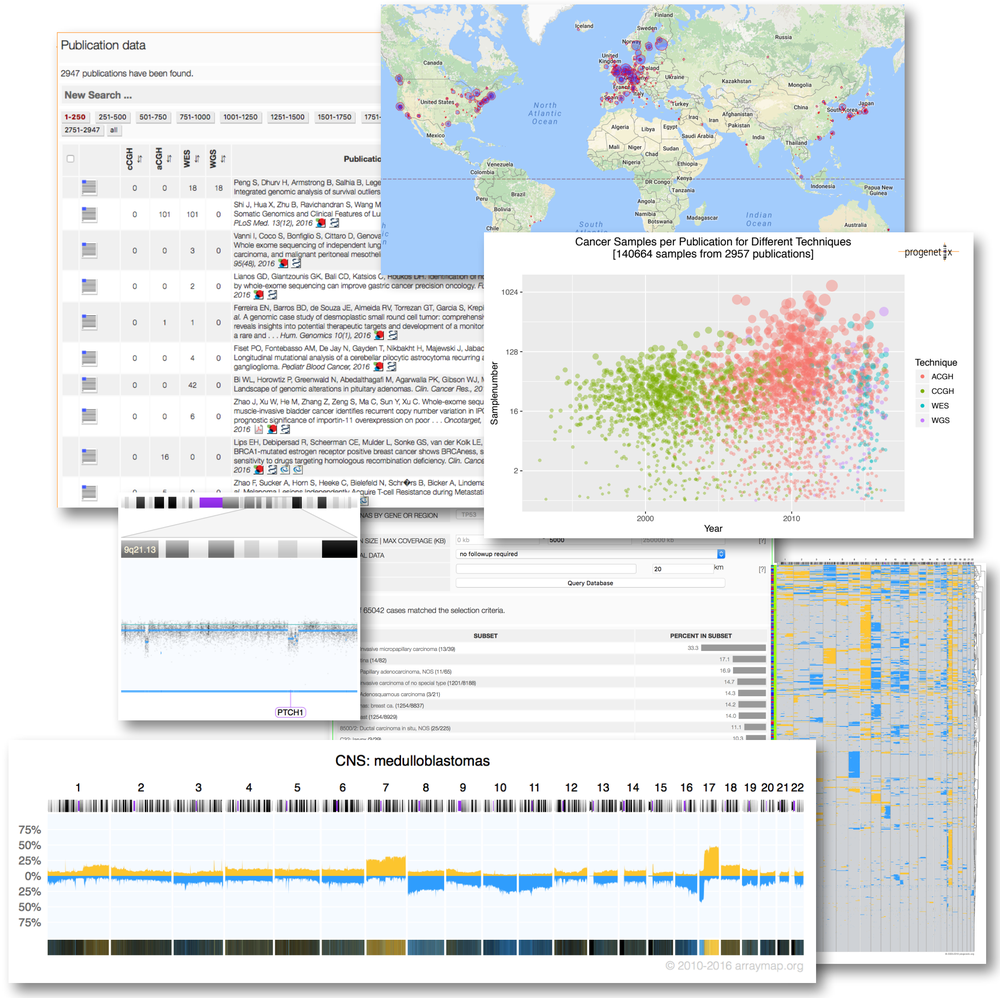
Our contributions I: Cancer genome knowledge resources

progenetix

- curated cancer genome publication resource (more than 3000 manually curated articles)
- article metadata and annotated genome profiles
- ontology mappings; clinical data where available
- epistemology: geographic and histologic sampling biases

arrayMap

- more than 60'000 array based genome profiles
- probe level, copy number, metadata
- completely open data access through web interface, downloads and API calls
- re-annotated metadata (diagnostic coding, basic clinical) for all samples



arrayMap

Resource for copy number variation data in cancer

arrayMap

- Search Samples
- Search Publications
- Gene CNA Frequencies
- User Data
- Array Visualization
- Progenetix

University of Zurich[UZH]

- Citation
- User Guide
- Registration & Licensing
- People
- External Links ↗

FOLLOW US ON twitter

(cc) BY-SA

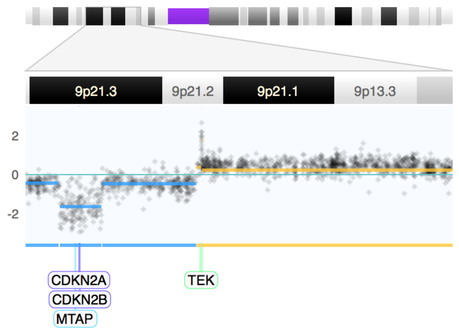
130.60.23.21

visualizing cancer genome array data @ arraymap.org

arrayMap is a curated reference database and bioinformatics resource targeting copy number profiling data in human cancer. The arrayMap database provides an entry point for meta-analysis and systems level data integration of high-resolution oncogenomic CNA data.

The current data reflects:

- 63060 genomic copy number arrays
- 763 experimental series
- 145 array platforms
- ICD-O 141 ICD-O cancer entities
- 554 publications (Pubmed entries)



Genomic copy number imbalances on chromosome 9 in a case of Glioblastoma (**GSM491153**), indcating, among others, a homozygous deletion involving CDKN2A/B.

ICD-O | Locus | ? | HG18 | HG19

For the majority of the samples, probe level visualization as well as customized data representation facilitate gene level and genome wide data review. Results from multi-case selections can be connected to downstream data analysis and visualization tools, as we provide through our Progenetix project.

arrayMap is developed by the group "Theoretical Cytogenetics and Oncogenomics" at the Institute of Molecular Life Sciences of the University of Zurich.

BRAIN TUMOURS	5653 samples ↗	[?]
BREAST CANCER	8329 samples ↗	[?]
COLORECTAL CANCER	3238 samples ↗	[?]
PROSTATE CANCER	991 samples ↗	[?]
STOMACH CANCER	1062 samples ↗	[?]

ARRAYMAP NEWS

2016-08-03: SVG graphics

2016-05-17: Transitioning to Europe PMC

More news ...

Feel free to use the data and tools for academic research projects and other applications. If more support and/or custom analysis is needed, please contact Michael Baudis regarding a collaborative project or a special license.

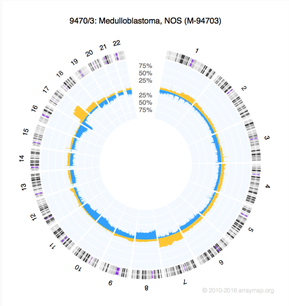
ICD Morphologies

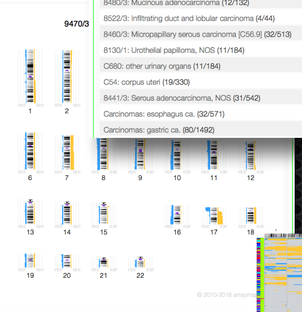
2021 samples from arraymap have an associated "ICDMORPHOLOGYCODE" label.

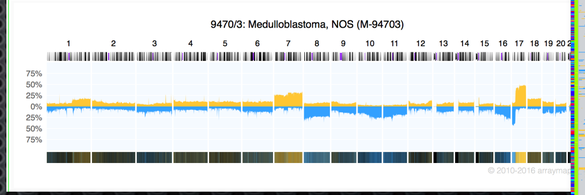
9470/3: Medulloblastoma, NOS (M-94703)

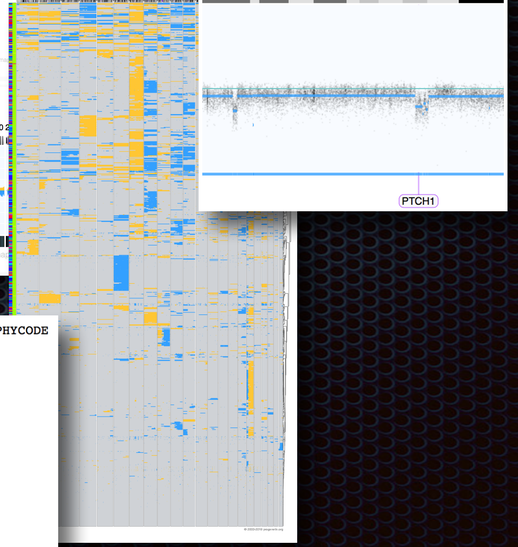
Synonyms

- Medulloblastoma, NOS
- Melanotic medulloblastoma









FIND CNAS BY GENE OR REGION: TP53 | [ERBB2] 17:35097862-35138441:1 [?]

REGION SIZE | MAX COVERAGE (KB): 0 kb - 5000 - 250000 kb [?]

CLINICAL DATA: no followup required

CITY: km [?]

Query Database

1949 of 65042 cases matched the selection criteria.

SUBSET	PERCENT IN SUBSET
8507/3: Invasive micropapillary carcinoma (13/39)	33.3
C692: retina (14/82)	17.1
8260/3: Papillary adenocarcinoma, NOS (11/65)	16.9
8500/3: Invasive carcinoma of no special type (1201/8188)	14.7
8560/3: Adenosquamous carcinoma (3/21)	14.3
Carcinomas: breast ca. (1254/8837)	14.2
C50: breast (1254/8929)	14.0
8500/2: Ductal carcinoma in situ, NOS (25/225)	11.1
C32: larynx (3/29)	10.3
8010/2: Carcinoma in situ, NOS (2/20)	10.0
C187: sigmoid incl. rectosigmoid junction (13/140)	9.3
8480/3: Mucinous adenocarcinoma (12/132)	9.1
8522/3: Infiltrating duct and lobular carcinoma (4/44)	9.1
8460/3: Micropapillary serous carcinoma [C56.9] (32/513)	6.2
8130/1: Urothelial papilloma, NOS (11/184)	6.0
C680: other urinary organs (11/184)	6.0
C54: corpus uteri (19/330)	5.8
8441/3: Serous adenocarcinoma, NOS (31/542)	5.7
Carcinomas: esophagus ca. (32/571)	5.6
Carcinomas: gastric ca. (80/1492)	5.4

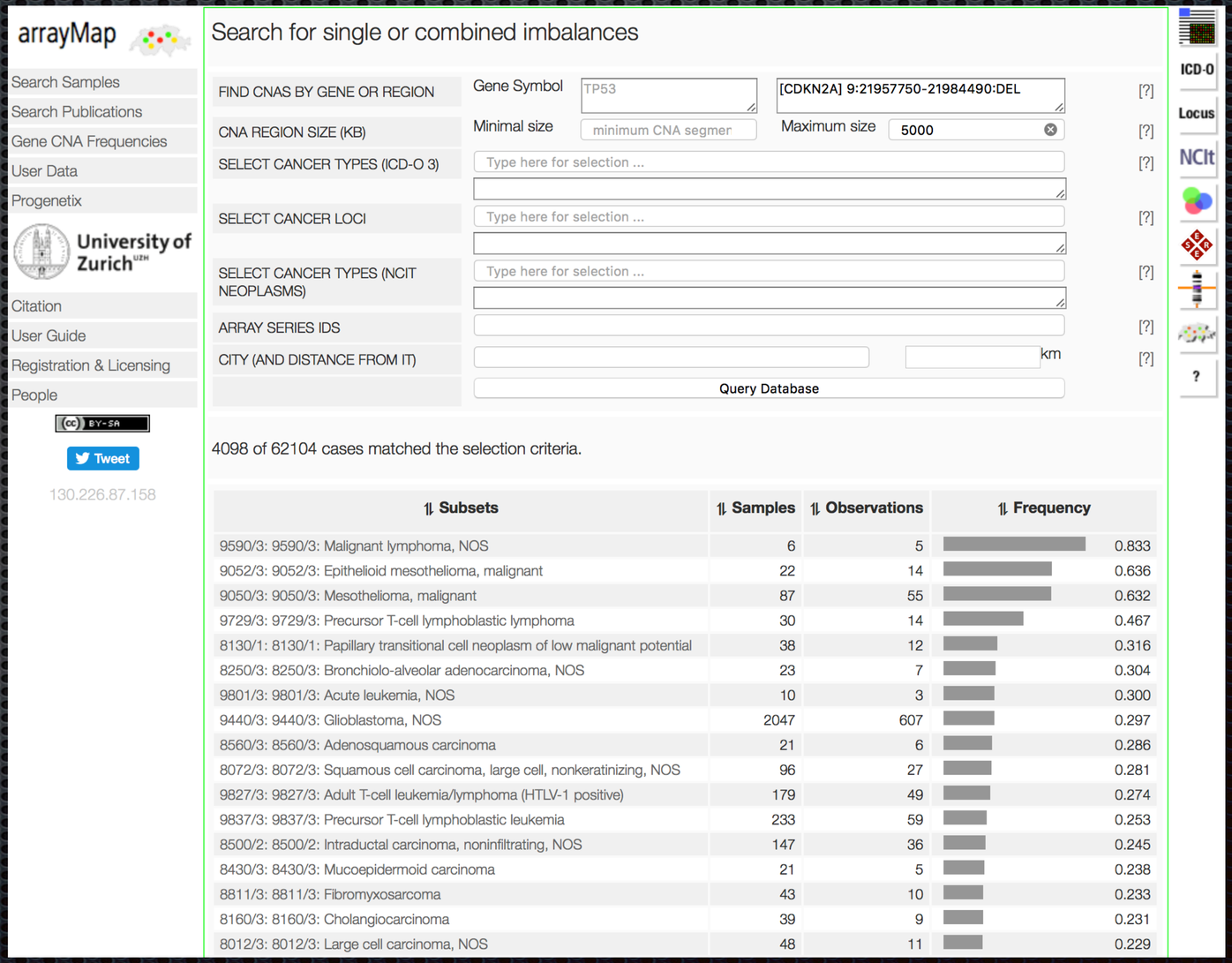
UID	SERIESID	PMID	ICDMORPHOLOGYCODE	ICDTOPOGRAPHYCODE
GSM1000061	GSE36942	23457519	8070/3	C10
GSM1000062	GSE36942	23457519	8070/3	C10
GSM1001316	GSE40777	23571474	8070/3	C53
GSM1001317	GSE40777	23571474	8010/3	C34
GSM1001318	GSE40777	23571474	8070/3	C09
GSM1001319	GSE40777	23571474	8010/3	C34
GSM1002668	GSE40834	24047479	9823/3	C42
GSM1002669	GSE40834	24047479	9823/3	C42
GSM1002670	GSE40834	24047479	9823/3	C42
GSM1002671	GSE40834	24047479	9823/3	C42
GSM1002672	GSE40834	24047479	9823/3	C42
GSM1002673	GSE40834	24047479	9823/3	C42
GSM1002674	GSE40834	24047479	9823/3	C42
GSM1002675	GSE40834	24047479	9823/3	C42
GSM1002676	GSE40834	24047479	9823/3	C42
GSM1002677	GSE40834	24047479	9823/3	C42
GSM1002678	GSE40834	24047479	9823/3	C42
GSM1002679	GSE40834	24047479	9823/3	C42
GSM1002680	GSE40834	24047479	9823/3	C42

arrayMap

arrayMap

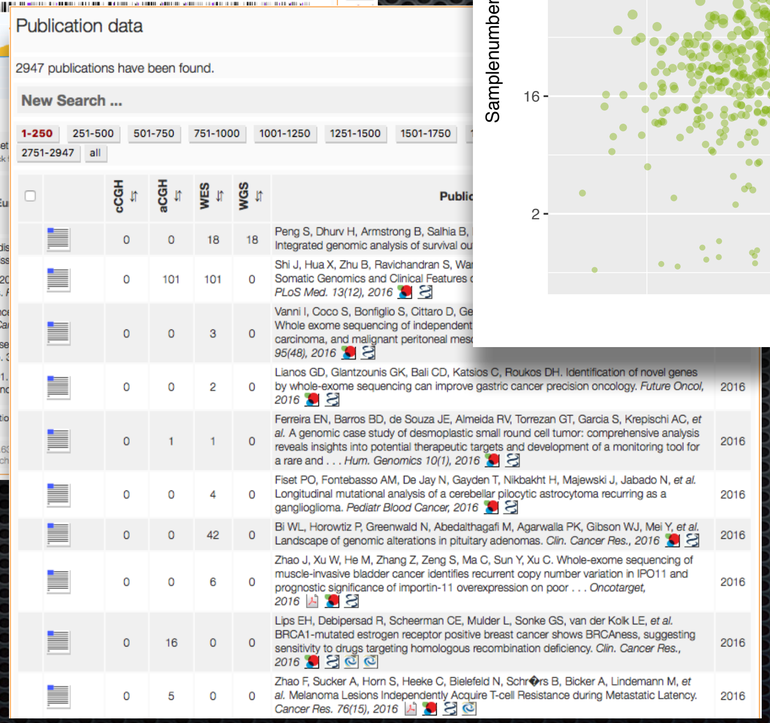
Resource for copy number variation data in cancer

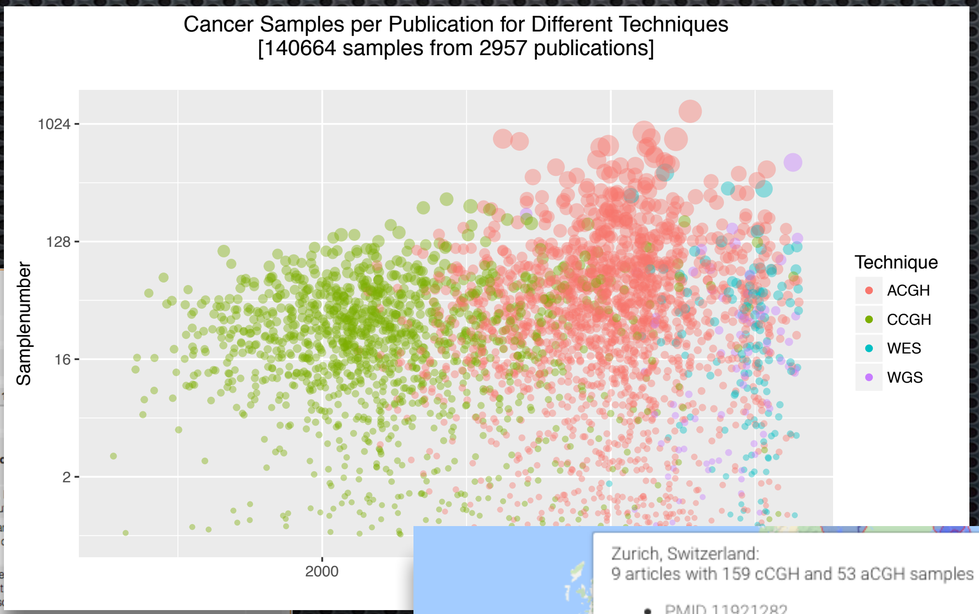
- Typical use: Gene CNA frequencies
- allows for (gene symbol supported) query for DUP/DEL overlapping a region
- size windowing ("focal" changes)
- no thorough statistics implemented yet (e.g. modeling overall CNA background)

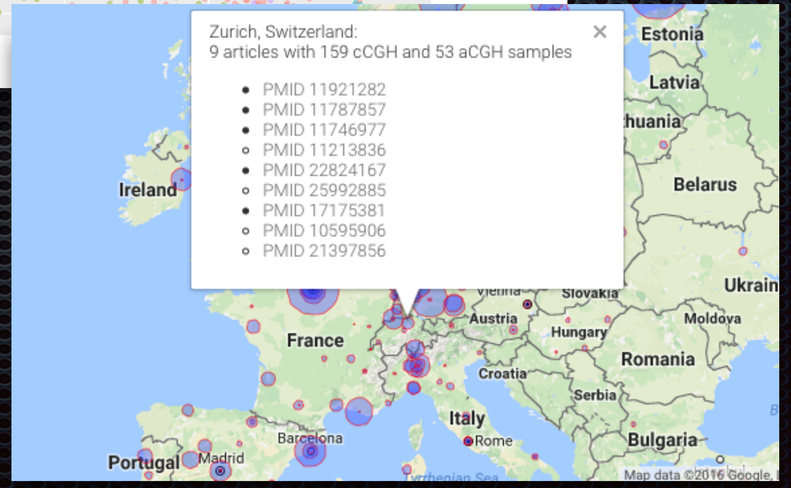


arrayMap

Progenetix: Cancer Genome Profiles, Article Metrics, Epistemology, **Resource Hub**



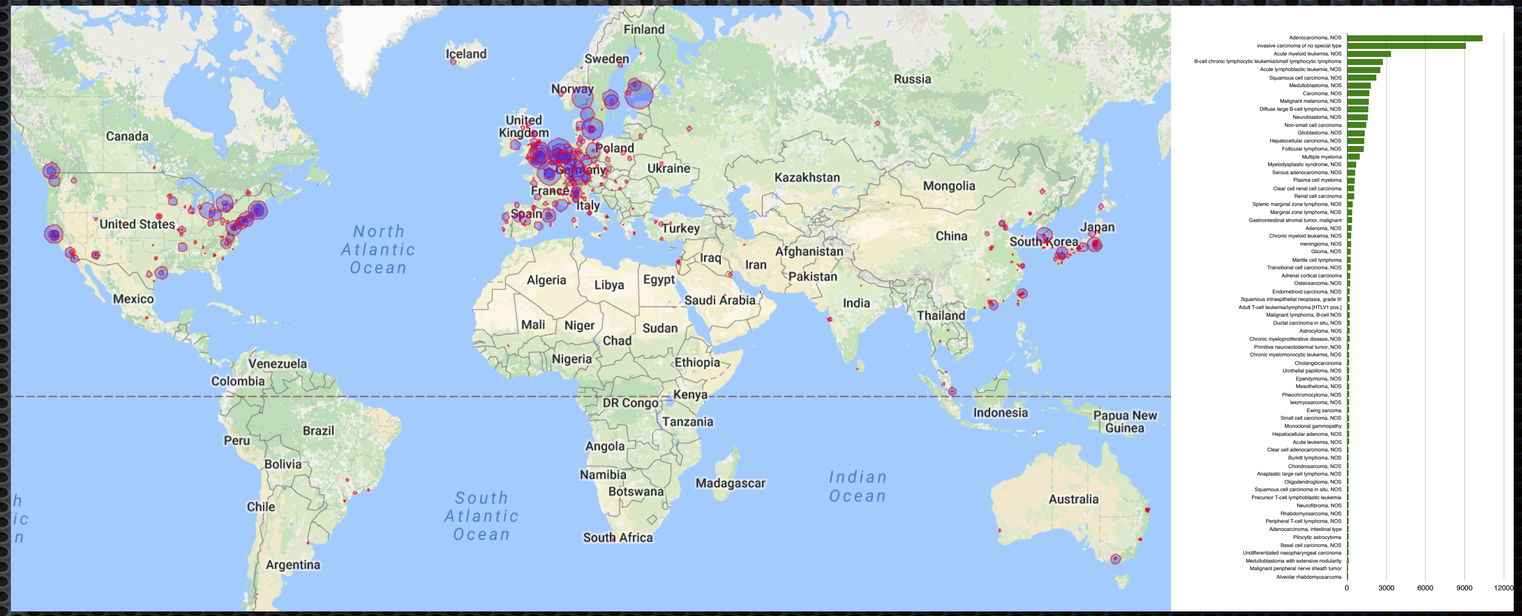




progenetix

Bias in Ascertainment / Background / Environment in Cancer Genome Studies

- the frequency of many genome variants depends on the genetic background
- cancer incidence & type can correlate to environmental factors
- geographic analysis can support interpretation and point to knowledge gaps



Geographic distribution of >140'000 cancer genome profiles reported in the literature. The numbers are derived from the 2947 publications registered in the Progenetix database.

progenetix

Population stratification in cancer samples based on SNP array data

- 2504 genome profiles from 1000 Genome project phase 1 as reference
- 5 superpopulations: South Asia, Europe, South America, East Asia and Africa.
- SNP positions used in 9 Affymetrix SNP arrays are extracted to train a population admixture model.

University of Zurich[UZH]

Prof. Dr. Michael Baudis
Institute of Molecular Life Sciences
University of Zurich
SIB I Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

arraymap.org

progenetix.org

info.baudisgroup.org

sib.swiss/baudis-michael

imls.uzh.ch/en/research/baudis

Global Alliance
for Genomics & Health